DOCUMENT RESUME

ED 297 585 FL 017 494

AUTHOR Spencer, Mary L.
TITLE Program Evaluation and Testing in the Educational Programs of Micronesia--When Benchmarks are Needed.
INSTITUTION Guam Univ., Mangilao. Coll. of Education.
SPONS AGENCY Office of Bilingual Education and Minority Languages Affairs (ED), Washington, DC.
PUB DATE Feb 85
NOTE 18p.; Paper presented at the Annual Meeting of the Pacific Island Bilingual Bicultural Association (Truk, Micronesia, February 1985).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Bilingual Education Programs; Chamorro; *Educational Assessment; *English (Second Language); Evaluation Criteria; Foreign Countries; *Language Tests; *Program Effectiveness; Program Evaluation; Second Language Programs; Summative Evaluation; Uncommonly Taught Languages
IDENTIFIERS Hawaii; *Micronesia

ABSTRACT

A discussion of bilingual education in the Pacific region focuses on the role of evaluation and testing in promoting educational quality. Based on a review of proposals, evaluations, and tests of all districts in Micronesia and Hawaii in the 1982-83 program year, these conclusions are drawn about program evaluation: (1) little program evaluation was undertaken; (2) student assessment was by far the most serious problem; (3) two evaluation design problems are occurring, the use of control group design and the use of summative rather than formative evaluation of new programs; (4) one seemingly good evaluation was seriously flawed by a lack of objectivity; (5) evaluation of Micronesian projects is underbudgeted; and (6) additional technical assistance for contract development and administration is needed. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Presented at the annual meeting of the Pacific Island Bilingual Bicultural Association, Truk, February, 1985.

PROGRAM EVALUATION AND TESTING IN THE EDUCATIONAL PROGRAMS OF MICRONESIA
- WHEN BENCHMARKS ARE NEEDED -

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
M. Spencer
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Mary L. Spencer, Ph.D.
Evaluation Specialist
Project B.E.A.M.
(Bilingual Education Assistance in Micronesia)

ED297585

When the U.S. Department of Education's Office of Bilingual Education and Minority Language Affairs contracted for the regional multifunctional service centers in 1983, it mandated that the centers provide training and technical assistance in program evaluation and documentation. This mandate was inspired by a history of evaluation neglect and inadequacy in the Title VII bilingual education programs (e.g., Okada et. al., 1982)

In my role as Evaluation Specialist for Project B.E.A.M. over the past year, I have worked on-site and in regional meetings to provide training on the fundamentals of evaluating bilingual educaton programs, and for testing or developing tests of primary languages and of English. The travel that the center's contract has made possible has permitted me to directly observe both Title VII and regular classrooms in many places in Micronesia, to review past evaluation reports and proposals, as well as five year plans for Yap and in CNMI. The LEAs in the Region have provided Project B.E.A.M. with all of the oral language and literacy tests used in the schools of Micronesia, which I have reviewed for linguistic and psychometric quality. And, it has become possible in some instances to work with educators who are currently developing new tests, or revising tests used in the past.

FL017494

So, I come to you today with one of my purposes being to share what I have seen and read and discussed - as a sort of condensation, from my point of view, of the state of testing and evaluation in the Region.

But this is only part of my purpose. I believe that program evaluation and student assessment must be useful as an integral part of our educational programs. They must begin at the same time that you begin planning your programs, and then continue as an interlocking part of our programs.

Of course, our Title VII bilingual education programs mandate that we evaluate them. When we accept Title VII money, we promise the government that we will evaluate our program. But frankly, if evaluation and testing are not truly useful as integral parts of our programs - we may see them simply as frills we cannot afford - and ultimately the program evaluation may not get done, or it may get done in an unsatisfactory way.

So, to explain the role of evaluation and testing as essential and integral to bilingual education in the Pacific, I must speak of more than evaluation designs, test-retest reliability, control groups, and significant differences. I must also speak about education in the Pacific.

Let us look for a moment at the word evaluation and try to see what its meaning is in terms of education. The word evaluation means the process of valuing or judging our educational programs. When we evaluate our educational programs we are trying to make statements about their merits and even about their deficiencies. We are describing what we have, and weighing the strengths and the weaknesses, and then forging new approaches that build on strength and reduce or eliminate the weak spots. But this process of valuing and judging and then moving ahead with improvements does not happen in a blank or sterile environment. It happens within the context of our ideas of what is good in education, particularly in bilingual education, and what is bad - the ideas and knowledge that we already have...and then of course these ideas grow and change as our documentation and evidence expands.

So, when we evaluate, we are looking for something against which to compare our situation (our staff development, parent involvement, material development, student achievement) so we know how it measures up; so we know whether our situation is good, medium, or bad. We do this same sort of thing in our everyday lives too. If I go into the lagoon and catch a certain kind of fish in my net for the first time, how will I know that it is a good one? How will I know whether it is a large sardine or just a sick puny tuna? If I am weaving a basket, how will I know that the design is good and that the shape and the texture have quality?

A few weeks ago I was in the Women's Handicraft Collective in Majuro looking at baskets and mats. I asked one of the Marshallese women there how I would know which mat was better than the others. She showed me several mats and pointed out the width of the pandanas strips that are used to weave the mats. She said, "Notice how the strips in this mat are thinner than in the other mats. That means that the maker had to work hard to make more strips and to make them just the same width; and notice how all of the strips are one color in these mats, but that this mat is better because it has strips of two different colors - one light and the other darker. And notice how the different colors are crossed in a pattern. The design in this fine mat is perfectly spaced so that the distance between the dark patterns is exactly the same in each block."

So, I will know a good fish, or a good basket, or a good mat by looking and comparing with other good examples...other fish, or baskets, or mats. And also by the consensus, the knowledge, the agreement from others who know about these things - from evidence and from expert judgments. I will call these things benchmarks.

Throughout this week's PIBBA Conference, we have had outstanding presentations and discussions on theories of language development, creative approaches to bilingual instruction, descriptions of system-wide planning for bilingual and general education curriculums, and recommendations on the schedules and concentrations of language use throughout the grades.

I think that we have all been listening carefully because we want to learn lessons from these presentations - we are searching in them to find the best evidence and the best expert judgments to guide us as we work to make the bilingual programs of the Pacific as good as they can possibly be. We are listening carefully because we are looking for benchmarks.

Remember for a moment what we heard. We heard about:

- o literacy studies on some non-western peoples - people in Liberia;
- o we heard a little about the effective schooling research done in the United States;
- o we heard about the experience of Finnish children who moved to Sweden and went to school there;
- o we heard about English speaking middle income Canadian children who went to schools where all of their instruction was in French;
- o we heard about other non-English speaking children who immigrated to Canada and went to school there;
- o we heard about some research done on LEP students (mostly Spanish speaking LEP students) who go to school in California and the way their teachers use language in their instruction.

And,

- o we heard about studies of education in Samoa 13 years ago compared to very recent times;
- o we heard expert judgments and discussions on language policy issues throughout the Pacific from highly qualified local Pacific authorities;
- o and we saw how David Ramarui's concepts were being used to guide the development of system-wide curriculum in Kosrae.

What I see us doing here is trying very hard to identify the best research information we can find to use as our benchmarks. But the problem we are having is that almost none of the evidence is on our own Pacific children and schools, and it does not speak to our own unique language environments. And this becomes an even more critical problem when we look around the room for our experts to obtain their judgments - and find that they often disagree. What do we do when we lack evidence and our experts disagree? Maybe we could have a duel - or maybe we could have them shout at each other, accepting the one who speaks louder or more often. Of course we cannot be silly. Too much is at stake. The well being and development of Pacific children and of their many Pacific nations are too serious for jokes. We need local evidence on the relationships that bind Pacific language, culture, and education. With local evidence we cxan develop better expert judgment and more consensus amongst our experts.

Let us look for just a moment at some of the critical concerns about Pacific bilingual education that were raised at this meeting; i.e., concerns that call for benchmarks - evidence that guides us toward better bilingual education programs. And, I will phrase these as broad evaluation questions that are important to the Pacific in general.

1. What is the relative feasibility and effectiveness of a transitional bilingual education model versus a maintenance model with children who enter school with a vernacular language as their primary language?

2. What is the feasibility of implementing a restoration model with Pacific children who enter school more proficient in English than in their vernacular language? And, if this is feasible, what are some of the most effective ways to implement this model?

3. Being careful to distinguish a bilingual education model (transition, maintenance, or restoration) from bilingual

teaching methods (such as preview-review, concurrent, translation, alternate day, or activity specific approaches to using each language of instruction), which methods of instruction work best in certain places, with certain kinds of resources, and for certain purposes? What is student achievement like with these different methods? When is it advisable to use each?

4. When is the best time to introduce English to Pacific children who enter school as monolingual vernacular speakers or who enter with very limited English proficiency?

   a. Should English be introduced early, and if so, should only oral English be introduced at this time, or should both oral English and literacy instruction be introduced at this time?
   b. Should English be introduced later, and if so, at which grade and in what way?
   c. Or should a dual language approach be used early. If both languages are used early, should children later be transitioned away from their primary languages, or should both the vernacular and the English language be maintained?

5. What is the effect of our diverse Pacific language environments on how well children do in English and vernacular language arts under different bilingual education models and instructional approaches? What are the implications of these language environments for the grade at which oral English instruction is introduced, or the time for beginning English reading and writing instruction, and for whether or not we must eliminate vernacular instruc-

tion in the upper grades in order to achieve English oral proficiency and literacy.

6. How many minutes or hours a day do we need to teach a language in order for a child to have Basic Interpersonal Communication Skills in that language? How does this vary in our different language environments?

7. How many minutes or hours a day do we need to teach a language in order for a child to have the complex thinking and literacy skills that are necessary for academic accomplishment in that language?

8. The South Pacific Commission and Tate Oral materials have been used in the Pacific for many years. And, they are used almost universally throughout the South Pacific and Micronesia. Weldis Weldy and Elizabeth Rechabai estimate that the SPC materials have been used in the Pacific for nearly 20 years. But their effectiveness has never been evaluated.

   What is the relative effectiveness of the Tate Oral Language Approach and the SPC materials compared to some of the newer ESL approaches such as the Total Physical Response or methods which emphasize natural use of oral language, deemphasize group reading aloud, or make use of basal and supplementary readers such as Laidlaw, Ginn, etc.

9. What is the oral English proficiency of Pacific children as measured by valid, reliable, comparable tests of oral English proficiency? And what effect do our bilingual education programs have on it after one, two, and three

years? For almost a decade, we have had bilingual education programs in the Pacific, each one of which has had English proficiency as a main objective. We have the tests, but we do not have the answer to this question in most places. We gave oral English data in Hawaii although it has not been organized in ways that allow us to answer this question. The Guam bilingual education and language programs have conducted a study of Faneyakan that has given us the answer to at least part of this question for Chamorro children in Guam. If each bilingual education program tested their students, or even a sample of their students this Spring, we could have resounding answers to at least part of this important question by the next time PIBBA meets.

10. Are the educational needs of girls and boys being equally well met by Pacific bilingual education programs? Are there gender differences in effective practices? Are there gender differences in enrollment, attendance, and achievement in different subjects or in different grades?

If we had answers - even partial answers - to some of these questions, we could use these as <u>benchmarks</u> to help resolve some of the concerns you have been raising:

1. Repeatedly during this conference, we have heard that one of the main constraints to providing effective vernacular language arts and content area education was the lack of materials.

   If we discovered that it is feasible to use a dual language approach in junior and senior highschool, and

that instead of choosing only one language that students will be literate in, that in fact they can become literate in two languages in the same amount of time, we could have our older students engaged in creative vernacular writing activities that could actually produce some of the vernacular materials that we need. And, at the same time develop the skills of our future Pacific authors, novelists, historians, poets, journalists, and writers of Pacific law.

2. We heard about the need to make parents and community members aware of what our programs are doing, and to give them evidence that we are fulfilling the goals that they have for their children's educations.

   The answers to these questions would allow us to do that.

3. We heard that language policy is hard to develop if schools and communities do not know what their ultimate goals are. I think that schools and communities are often not clear about what they want because, first, they are not sure what is possible under the best of all conditions; and second, because they are not sure what the resources of their own schools and communities can buy.

   If we had the answers to some of these evaluation questions, we could give them some of the benchmarks they are looking for.

So the question now is, "How are Pacific bilingual education programs doing in their program evaluations?" Some of the questions we have listed can only be

answered when we have evidence from a number of programs. But some of the questions (e.g., English oral proficiency) can be answered annually when individual programs simply document how they progressed toward their objectives.

Having reviewed the proposals, evaluations, and tests of all LEAs in Micronesia and Hawaii, I can give some fairly general answers about the status of program evaluations in the Pacific, based on the 1982-83 program year. Some evaluation activities are still in progress for the 1983-84 academic year, thus postponing consideration of the most recently completed school year.

1. In Micronesia, there are eight LEAs. In 1982-83, two of them had institutionalized bilingual education programs, and six of them had Title VII basic bilingual education programs. Neither of the institutionalized programs conducted evaluations per se, although both conducted various monitoring, documentation, and testing activities. And in the case of one of these, a comprehensive test of primary oral language proficiency was developed and field tested. Of the six Title VII programs, I would say that about 3 1/2 to 3 3/4 evaluation reports were prepared. The two programs represented by these fractions prepared documentation and testing information, but lacked narrative description and interpretation organized into a final report format.

   For the 1982-83 school year in Hawaii, an evaluation study was undertaken for one of the three basic projects. The following year, evaluation reports were developed for all three projects.

2. Testing of program students was by far the most serious evaluation problem - as it is for all of the Title VII

programs that I have evaluated in the United States.

Micronesia

a. Regarding the selection and use of English oral language proficiency tests, only Guam has documented the English oral language proficiency of their bilingual education student population with a test that has established and adequate validity and reliability information. There are two main problems associated with English oral language testing that affect almost all LEAs in Micronesia. First, the tests currently in use, where a test is used at all to document oral English proficiency, do not test for speaking ability. Since all of our programs claim that English speaking ability is one of their major objectives, the lack of such data is a major gap. Second, the programs have not established baselines for the oral English proficiency of their program students. By this, I mean that they need to test the children in the first few weeks of school during the first program year, in order to determine how proficient they were before the program began to increase their skills. It is essential to have baseline data in order to know how well students make progress toward the goal of becoming proficient in oral English language skills.

b. Regarding English reading tests, we generally find that either the Micronesian Achievement Tests Series alone, or MATS in combination with local reading items are being used. The MATS was developed for the purpose of assessing and comparing the status of English reading, math, and listening comprehension

skills of the LEAs throughout the Trust Territory. The MATS was seen as a normed test only in the sense of this very broad-gauged documentation and comparison purpose. However, it is now being used as though it were a truly norm-referenced achievement test for both individual students and groups of students. There are two main problems with the use of the MATS. First, when we look closely at the scores of students in most places, we see that there are many students who are obtaining perfect scores. It appears that the test is too easy, and therefore that many of the more advanced skills of the students cannot be tested by it. This causes the average of the test to be distorted. The average score will be lower than it really should be because many students cannot score as high as they are able to score. And this "topping out" effect causes other distortions in the results. Second, it is very hard to know what the scores on the MATS mean. We need a way of knowing what scores should be expected of students in a certain grade early or later in the school year. We need better benchmarks for English reading skills, particularly in the area of English reading comprehension with different types of reading purposes and materials.

c. Regarding oral vernacular proficiency, we find that many of the Micronesian LEAs use a small number of test items to assess basic levels of oral vocabulary listening comprehension. Yap, Ponape, Kosrae, and Palau have done this. The LEAs in the Marianas Islands have done this more extensively, as we would expect in view of their concerns for language pre-

servation and restoration. For example, Guam has developed a comprehensive test of listening and speaking for Chamorro children in Guam. In the Commonwealth of the Northern Marianas Islands, extended competency tests have been developed in Chamorro and also in Carolinian.

d. Vernacular reading tests are being developed or extensively modified throughout the Micronesia Region. The Chamorro and Carolinian competency tests in CNMI each contain a mixture of reading vocabulary, sentence and paragraph comprehension, and various forms of detail identification. In the 1982-83 school year, Yap used vernacular reading items of this type. Kosrae has such a test for fifth grade students, is currently developing a version of this that is suitable for sixth grade students, and plans to develop a system-wide vernacular reading test series. Palau has been working on individually administered reading comprehension measures. The Republic of the Marshall Islands is developing a vernacular reading test that will be included in the high school admissions test. A preliminary version has been developed and it will soon be field tested.

Hawaii

a. Regarding oral English proficiency, Hawaii has a sophisticated system for assessing English oral proficiency of language minority students. This system is used to identify students who are eligible for bilingual or ESL services. However, this data has not been organized in a manner that permits us

to understand what the baseline performance of bilingual education program students is, nor to show how much progress in oral English proficiency they make at the end of one, two, or three program years.

b. Regarding vernacular oral language testing, Hawaii uses an interview-type process for determining several levels of vernacular oral proficiency as part of the student identification and entry placement process.

c. The two secondary level Hawaii Title VII projects use a high school competency test, HSTEC, which assesses 15 different basic skills in English. Other measures of reading achievement in English are not used. The preschool Title VII project uses a test which combines conceptual and language aspects of performance.

d. The Hawaii projects do not measure vernacular reading skills.

3. Two evaluation design problems are occurring.

a. First, some people have tried to use a control group design. It is very hard to select a truly appropriate control group anywhere except in very large school districts. Even then there are ethical problems because the control group will be denied the educational opportunities that will be given to the program group. So when people have tried using control groups they often find that the control group is not truly comparable to the program group, and the comparisons must be discarded. In some cases, it is not possible to draw any conclusions because

of the way the data has been organized.

b. In Micronesia the programs have often been built one grade at a time. For example, the first program year of the grant might be spent developing a bilingual program in the first grade, then the next year the focus will be on the second grade, the third year on the third grade, and so on. This works fine for program implementation, but program evaluation cannot follow this same pattern. We all know that the first year that a program is implemented is its most difficult year. So we must follow students for more than just the first year that they are in the program. So far, this is not being done.

   The need for evaluation designs that follow the same program students for more than one year also applies to the Hawaii projects.

4. One project evaluation, although it had a very finished look, and was comprehensive in its coverage, was fatally flawed by its lack of objectivity. In many places, the descriptions and interpretations appeared to be highly personal attacks on a particular member of the project staff, and the evaluator appeared to have a serious conflict of interest.

5. I think that all or almost all of the Micronesian Title VII projects are underbudgeting their evaluations. I recommend that in your new proposals, that a minimum of $5,000 be requested for evaluation. This would provide sufficient funds for an external evaluator to fully analyze oral language and reading scores in both

English and the vernacular languages, as well as address the other program components. It would probably be necessary to have additional funds for travel and per diem if an external evaluator is to be used. I also recommend that the programs consider paying for part time testers so that the main program staff can concentrate completely on program implementation while testing responsibilities are given the full attention of a trained testing staff, even if this staff is only temporary.

6. Finally, I think it is probably true for bilingual education programs throughout the Pacific - and perhaps throughout the U.S. as well - that some technical assistance is needed to help project directors and LEAs handle the nuts and bolts business details of developing and administering contracts with external evaluators, in scheduling their payments and in actually getting them paid at appropriate times.

Let us draw this presentation to a conclusion with a story about evaluation in Kosrae that Elmer Asher tells. Elmer told me that when he was about 11 years old, shortly after the end of the war, a man visited each village in Kosrae, sending the message everywhere that all of the boys like Elmer must go to a particular building for an important meeting. Once the boys had gathered, the man drew a long line across the wall and ordered the boys to line up against the wall. Then he quickly singled out the boys who stood taller than the line, and with a sweeping gesture commanded, "Everybody above the line goes to school in Ponape."

That line on the wall in Kosrae was a benchmark in Pacific education in 1945. I think we were only lucky that Elmer stood above the mark. But in 1985,

40 years later, we deserve better benchmarks - better evidence on what makes our programs effective so that you - the experts on Pacific education can satisfy some of your questions and thereby improve your abilities to develop language policy and to make the best choices needed to guide Pacific education. And many of the answers we seek are within reach if only we keep our promises to conduct good solid evaluations of our programs. So I will look forward to PIBBA 1986 when some of you will surely be able to share new benchmarks for Pacific education with the rest of us.